PART4　データの解釈２

目標
- 単回帰分析や仮説検定の考え方について理解する。
- 単回帰分析や仮説検定の考え方を利用したデータ分析について考えられるようになる。

Chapter1　データの解釈２

単回帰分析

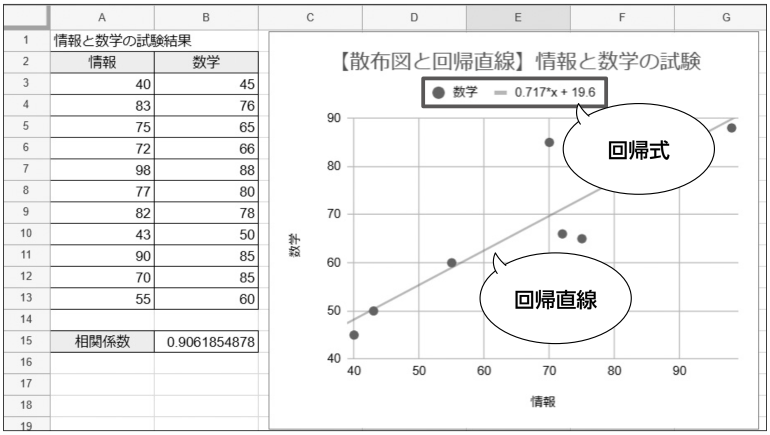
・単回帰分析と回帰式

2 つの事柄の関係性を数式の形で表す手法で，1 つの従属変数（Y）を 1 つの独立変数（X）で予測するものを〔①　　　　〕という。

散布図において，予測値を求める際に用いられる直線のことを〔②　　　　　　　〕といい，2 組のデータの中心的な分布傾向をあらわすものである

回帰直線は，Y = aX+b の〔③　　　　　〕で表すことができる。

例えば，次の図で情報の点数（X）と数学の点数（Y）に相関関係ある場合，Y = 0.717X+19.6 が回帰式となる。


情報と数学の試験結果

情報	数学
40	45
83	76
75	65
72	66
98	88
77	80
82	78
43	50
90	85
70	85
55	60
相関係数	0.9061854878

仮に，情報の点数（X）が 75 の場合，数学の点数（Y）=73.375 のように予測をすることが可能になる。

覚えておこう

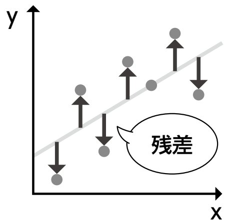
独立変数は，原因となる変数のことで，従属変数は結果となる変数のことである。
Y = aX+b の回帰式の場合，X が独立変数，Y が従属変数となる。
独立変数を説明変数，従属変数を目的変数と呼ぶこともある。

知っておこう

回帰分析には，単回帰分析のほかに重回帰分析がある。単回帰分析が，1 つの従属変数を 1 つの独立変数で予測したのに対し，重回帰分析は 1 つの従属変数を複数の独立変数で予測するものである。

- **最小二乗法**

　データの各点と回帰直線のＹ軸方向のずれがトータルで最も小さくなるように回帰直線を描く手法のことを〔④　　　〕という。図の場合，矢印の誤差（残差）が最小となる関係式を求める方法となる。

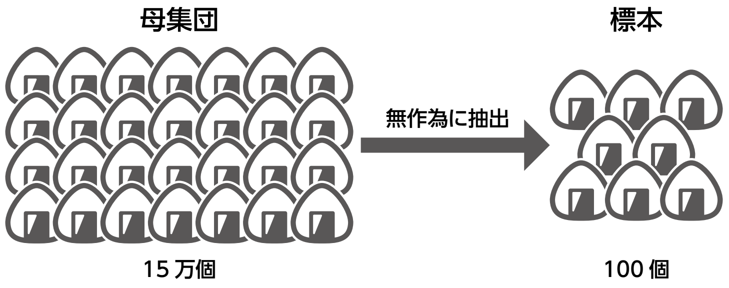


仮説検定の考え方

- **母集団と標本**

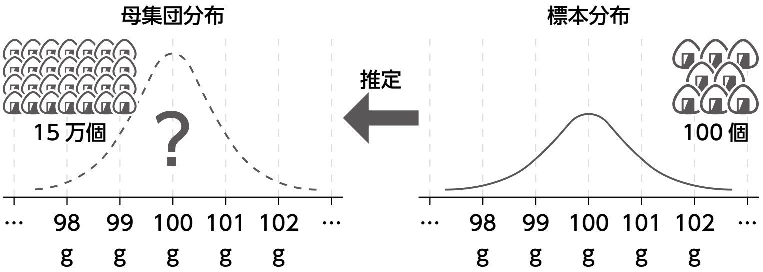
　統計対象となる全ての集合のことを〔⑤　　　　　〕という。

　母集団から抽出した部分集合のことを〔⑥　　　　〕という。



- **推定**

　標本をもとにして調査を行い，母集団がどのような値を持っているかを確率的に推測することを〔⑦　　　　〕という。統計学の世界では，無作為に抽出した標本を調査して，全体（母集団）を推測するという方法が用いられる。

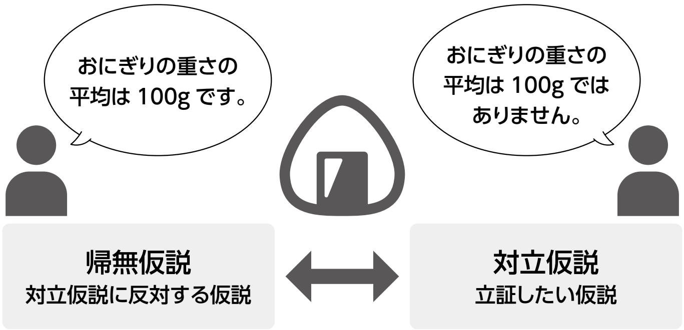


- **仮説検定の考え方**

標本のデータから，母集団に関する仮説の真偽を判断する手法のことを〔⑧　　　　　　　〕という。

例えば，1日15万個のおにぎりを製造する工場では，「おにぎりの重さの平均は100 gである」としている。本当に平均が100 gかを知るため，無作為に100個抽出して平均を調べた結果，標本の平均は104 gとだったとする。この結果だけから言えることは，「おにぎりの重さの平均は100 gではない」ということになる。しかし，工場長が「おにぎりの重さの平均は100 gのはず！標本の重さの平均が104 gとなったのは偶然です！」と言ってきたとする。

この時，「おにぎりの重さの平均は100 gではない」という立証したい仮説を〔⑨　　　　　　　〕といい，対立仮説に反対する仮説を〔⑩　　　　　　　〕という。



仮説検定では，帰無仮説が正しいという前提に立って，「対立仮説の結果はどれくらいの確率で起こり得るのか」を計算する。

この確率のことを〔⑪　　　　　〕（有意確率）という。統計学の世界ではp値が5%未満だった場合は，めったに起こらないことが起きたとして，対立仮説が正しいと判断する。このことを，帰無仮説を〔⑫　　　　　〕するという。

偶然ではなく必然である可能性があると推測されることを〔⑬　　　　　〕があるといい，この5%の基準を有意水準という。医療問題など，慎重な判断が必要な場合は，有意水準は1%が使われることもある。

重要

標本は全量ではないので，母集団と標本との間には誤差が生じる場合がある。この違いを**標本誤差**という。標本の数が少ないほど標本誤差は大きく，標本の数が多いほど標本誤差は小さくなる傾向がある。

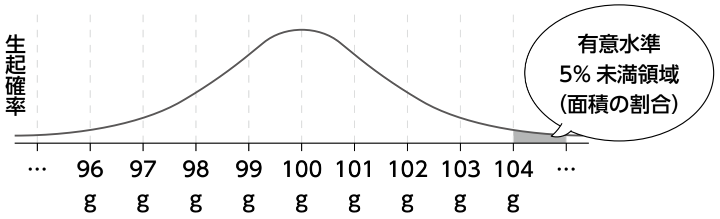
知っておこう

推定には，母平均＝100gのように1点で表す点推定，98g≦母平均≦102gのように一定の幅を持たせる区間推定といった種類がある。

知っておこう

同じ標本データで，都合よいp値が得られるまで何度も標本調査を行うことは，p値ハッキングまたはデータ・フィッシングと言われ，適切でない行動と考えられている。

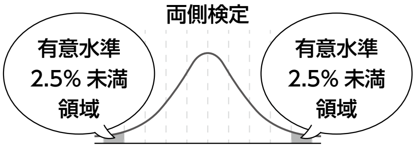
図は「おにぎりの重さの平均は100gである」とした場合の分布で，今回の標本調査の結果のように平均が104gとなるのは有意水準5%未満の領域であるから，前提となっている「おにぎりの重さの平均は100gである」という帰無仮説を棄却することになる。ここでの「棄却」とは，仮説を退ける，すなわち採用しないことをいう。帰無仮説を棄却したことにより，立証したかった対立仮説「おにぎりの重さの平均は100 gではない」が採択される。

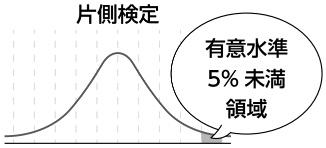


- **片側検定と両側検定**

大小の一方だけを検証する検定を〔⑭　　　　　　〕という。

大小の両方を検証する検定を〔⑮　　　　　　〕という。同じ有意水準5%でも，2つの検定で帰無仮説を棄却する範囲（棄却域）は異なる。両側検定は左右の領域を合わせて5%で考える。





まとめ

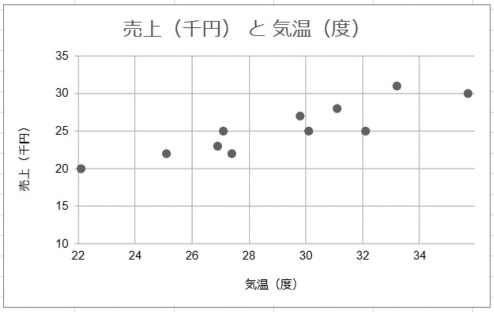
- **単回帰分析**とは，1つの従属変数(Y)を1つの独立変数(X)で予測するものである。
- **最小二乗法**とは，各点とY軸方向のずれがトータルで最も小さくなるように回帰直線を描く手法のことである。
- **仮説検定**とは，標本のデータから母集団に関する仮説の真偽を判断する手法のことである。
- 立証したい仮説を**対立仮説**といい，対立仮説に反対する仮説を**帰無仮説**という。

Chapter2　実践問題

問 1　次の図は，気温とアイスの売上の関係を表および散布図で表したものである。

(1)　表計算ソフトウェアの機能を使って，散布図に回帰直線と回帰式を表示させなさい。

	A	B
1	気温とアイスの売上	
2	気温（度）	売上（千円）
3	32.1	25
4	25.1	22
5	35.7	30
6	30.1	25
7	27.4	22
8	22.1	20
9	33.2	31
10	27.1	25
11	31.1	28
12	29.8	27
13	26.9	23
14		
15	相関係数	0.8992210763
16		



(2)　(1) の回帰式を利用して，気温が 37 度のときのアイスの売上を予測しなさい。単位は円とし，1 円単位で答えること。

問 2　次のア～ウの説明について，正しければ○，誤っていれば×で答えなさい。

ア　商品の重さの標本調査を行うため，母集団の中から重そうなものを抽出した。

イ　標本と母集団のデータの平均値は必ず一致する。

ウ　仮説検定では，帰無仮説が誤っているという前提に立って検証を行う。

問3 仮説検定を行う場合の手順について，以下の①～④を正しい順番に並べ替えなさい。

前提：ある牛丼屋では，肉の量が平均 90 g になるように規定している。無作為に 10 個重さを調べたところ平均は 95 g であった。

① 帰無仮説が正しいという前提のもと，平均が 95 g になる確率（p 値）を求める。

② 「牛丼の肉の重さの平均は 90 g である」という帰無仮説を立てる。

③ 計算の結果 p 値が 2%となったため，有意性があると判断し，帰無仮説を棄却する。

④ 「牛丼の肉の重さの平均は 90 g ではない」という仮説を立てる。

第9講 PART4　確認問題

問 1　1 つの従属変数 (目的変数) を, 1 つの独立変数 (説明変数) で予測することを（　ア　）という。係数 a，b を使ってこの 2 変数の関係を Y=aX+b という式で表すとき，この式を（　イ　）, この式で表される直線を（　ウ　）という。

(1) 空欄（ア）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

① 回帰式　② 回帰直線

③ 重回帰分析　④ 単回帰分析

(2) 空欄（イ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

① 回帰式　② 回帰直線

③ 重回帰分析　④ 単回帰分析

(3) 空欄（ウ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

① 回帰式　② 回帰直線

③ 重回帰分析　④ 単回帰分析

問 2　最小二乗法について述べた文章として最も適切なものを，次の①～④のうちから選びなさい。

① 2 つ以上のカテゴリ間でデータを比較するための集計方法

② データの各点と回帰直線の Y 軸方向のずれがトータルで最も小さくなるように回帰直線を描く手法

③ 元本に生じた利子を次期の元本に組み入れる方式

④ 複数の説明変数で 1 つの目的変数を予測する手法

問３　統計対象となる全ての集合のことを母集団といい，そこから抽出した部分集合のことを（　ア　）という。また，（　ア　）のデータから，母集団に関する仮説の真偽を判断する手法を（　イ　）という。このとき，立証したい仮説を（　ウ　），これに反対する仮説を（　エ　）という。（　エ　）が正しいという前提で（　ウ　）が起こり得る確率（p 値）を求め，一定値以下の場合は（　ウ　）がめったに起こらないとして，（　ウ　）が正しいと判断する。

(1)　空欄（ア）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　仮説検定　　②　帰無仮説
③　対立仮説　　④　標本

(2)　空欄（イ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　仮説検定　　②　帰無仮説
③　対立仮説　　④　標本

(3)　空欄（ウ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　仮説検定　　②　帰無仮説
③　対立仮説　　④　標本

(4)　空欄（エ）に入る語句として最も適切なものを，次の①～④のうちから一つ選びなさい。

①　仮説検定　　②　帰無仮説
③　対立仮説　　④　標本

第９講　データの活用

PART1　データの収集と整理

問１　③→①→②

問２　（ア）④　（イ）①　（ウ）③　（エ）②

問３　（ア）○　（イ）○　（ウ）×　（エ）○

問４　エ

PART2　データの分析

問１　（ア）×　（イ）○　（ウ）×　（エ）×

問２　2019 年に比べて 2020 年の方が最大値，最小値，平均値いずれも高く，暑い日が続いたと考えられる。

問３　省略

問４　269，サッカー，女子

PART3　データの解釈１

問１　(1) 省略　(2) 省略

(3) 散布図が右肩上がり，相関係数が 0.9 以上であることから強い正の相関があるといえる。

問２　（ア）×　（イ）×　（ウ）○

PART4　データの解釈２

問１　(1) 省略　(2) 31577 円

問２　（ア）×　（イ）×　（ウ）×

問３　④→②→①→③

第 9 講　PART4　確認問題　解答

問 1

(1)　④

解説：

1 つの従属変数を 1 つの独立変数で予測することを単回帰分析という。

(2)　①

解説：

この式を回帰式という。

(3)　②

解説：

この式で表される直線を回帰直線という。

問 2　②

解説：

最小二乗法は，直線とデータの y 方向の差（残差）の和が最小となる直線を求める手法である。

なお，①はクロス集計，③は複利法，④は重回帰分析の説明である。

問 3

(1)　④

解説：

母集団から部分的に取り出したデータを標本という。

(2)　①

解説：

母集団の仮説の真偽を判定する手法を仮説検定という。

(3)　③

解説：

仮説検定で立証したい仮説を対立仮説という。

(4)　②

解説：

立証したい仮説に反対する仮説を帰無仮説という。